

University of Dundee

## Face-to-Face Peer Assessment in Teacher Education/Training

Topping, Keith James

*Published in:*  
The Educational Review, USA

*DOI:*  
[10.26855/er.2021.05.002](https://doi.org/10.26855/er.2021.05.002)

*Publication date:*  
2021

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

*Citation for published version (APA):*  
Topping, K. J. (2021). Face-to-Face Peer Assessment in Teacher Education/Training: A Review. *The Educational Review, USA*, 5(5), 117-130. <https://doi.org/10.26855/er.2021.05.002>

### General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Face-to-Face Peer Assessment in Teacher Education/Training: A Review

Keith James Topping

University of Dundee, Scotland, UK.

**How to cite this paper:** Keith James Topping. (2021). Face-to-Face Peer Assessment in Teacher Education/Training: A Review. *The Educational Review, USA*, 5(5), 117-130.  
DOI: 10.26855/er.2021.05.002

**Received:** March 30, 2021

**Accepted:** April 27, 2021

**Published:** May 19, 2021

**Corresponding author:** Keith James Topping, University of Dundee, Scotland, UK.

**Email:** k.j.topping@dundee.ac.uk

## Abstract

This is a review of research literature on peer assessment (PA) on a face-to-face basis (without the use of technology) in teacher education and training (including pre-service and in-service teachers). PA is not just a tool for managing teaching burdens, but a mechanism for enhanced learning. Studies show PA has a consistent moderate effect and is at least as effective as assessment by teachers. A large number of databases were searched, from 2000-2020. Seventy-three papers were included. Participant perceptions of PA were reviewed, followed by studies of more objective PA outcomes. Most studies found that PA was effective and well received, but sometimes there was initial student resistance. Evidence-based design considerations for PA were then given. Limitations of the research were reviewed and the implications for practitioners, researchers and policy-makers outlined. Overall, peer assessment seemed an effective pedagogical tool for teachers, but there was little evidence of experience of it carrying over into use with school students.

## Keywords

Peer Assessment, Review of Research, Pre-Service, In-Service, Effectiveness

## A. Introduction

This paper is a review of peer assessment (PA) in teacher education and training, intended to explore outcomes but also extract points for designing future PA interventions better. PA in teacher education and training is an interesting area to explore, as one might expect teachers engaging in PA themselves would be more likely to initiate PA interventions for their students and also design these PA interventions more carefully. This review explores whether this is the case and is also the first review to explore this area.

### 1. Definition and history

A definition of peer assessment (PA) could be: "An arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners" (Topping, 2018). Specification could be in terms of marks, grades or scores, and/or more richly in terms of qualitative feedback. Feedback has, of course, been found to be highly effective in enhancing subsequent performance (e.g., Hattie & Wollenschläger, 2014).

PA is not just a tool for managing teaching burdens, but a mechanism for enhanced learning (Topping, 2018). For the assessor, the intellectual demands of reflecting, making a balanced assessment, and formulating and delivering the feedback, all lead to learning gains. For the assessee, the intellectual demands of receiving and evaluating the feedback, deciding what aspects to implement and what not, and reflecting on other issues prompted by the feedback but not contained within it, all lead to learning gains. Where the assessor and assessee have a further conversation to negotiate an

agreed compromise for onward progression, even further learning gains can be expected. Thus, both assessor and assessee are expected to gain from the activity. These mechanisms have been shown by research over many years (e.g., Annis, 1983). PA is not of course new. For example, George Jardine, Professor at the University of Glasgow 1774-1827, described a pedagogical plan including methods, rules and advantages of PA of writing (University of Glasgow Story, 2020).

## 2. Research evidence, reliability and validity of PA in general

Reviews and meta-analyses tend to be about higher education in general (Kane & Lawler, 1978; Topping, 1998; Falchikov & Goldfinch, 2000, Van Gennip, Segers, & Tillema, 2009; Hodgson, 2010; Tillema, Leenknecht, & Segers, 2011; Hoogeveen & van Gelderen, 2013; Li et al., 2016.; under review 2020), but not specifically about teacher education and training.

Aggregating the findings from these studies, it seems:

- PA has an effect size of 0.55-0.63 cf. no assessment (i.e., a consistent moderate effect),
- PA is similarly effective to assessment by teachers,
- There are similar effects at primary school, secondary school, and higher education.

Studies often assume that staff or “expert” assessment is the benchmark criterion for reliability and validity; but (as noted above) studies of peer marking/grading show it is as reliable as staff assessment; although staff assessment is not very reliable (Topping, 2018).

## 3. Methodology for this study

Since we were interested not only in outcomes but also in relevant underpinning theory and practical implications for the design of future programmes, the research questions were:

1. What coherent theory is visible in the research literature on PA in teacher education and training (TET)?
2. What empirical data on outcomes are visible in the research literature on PA in TET?
3. What evidence-based design principles are visible in the research literature on PA in TET?

The search focused only on papers in the English language. Search Terms were: “peer assessment” AND “teacher education” OR “teacher training”. Databases searched were: ERIC, APA PsychNet, JSTOR, ScienceDirect, the Social Science Research Network, CORE, ResearchGate, the Directory of Open Access Journals, EBSCO Teacher Reference Centre, JURN, Semantic Scholar, Paperity and Google Scholar. Given that PA is a relatively new topic in TET, it was decided to limit the search to 2000 or after. Five items turned up dated before 2000, however, and were included as they met the inclusion criteria. Given that a large number of databases were searched and the search keywords were very broad, there was no manual checking of citations in the relevant publications found.

The Criterion for Inclusion in the review was that the item addressed the research questions. It was expected that all selected papers would address the second research question and some would also address the first and/or third research questions. An exclusion criterion was that studies should not address PA with any form of technology, since PA with technology was to be the focus of a separate investigation.

In total 73 papers met the criterion for inclusion. Most of the items (n=64) were peer reviewed journal articles, but there was also a book, three book chapters, three conference papers, a doctoral thesis and an online reference. In this review, the terms “student” and “teacher” are easily confused. Generally, “pre-service teachers” means teachers in pre-service training, “in-service teachers” means teachers already in teaching posts (whether qualified or not) and receiving further training. Both may be referred to as “students”, but “students in schools” means children in a school classroom. “Teachers” means teacher educators or others with responsibility for managing and assessing student work. As this paper progresses, papers dealing with pre-service and in-service teachers are dealt with separately, since there is clearly considerable difference in the contexts in which they are operating, and findings from one area might not readily transfer to the other.

## B. Findings

### 4. Theory of PA in TET

There is a good deal of theory about PA in general (e.g. see Topping, 2018, chapter 4), but here we focus only on theory in PA in TET, and there was very little of it. Only one paper, by Sluijsmans and Prins (2006), touched on theory. They noted that PA among teachers involved collaboration, communication, discussion, critical reflection and analysis, and reciprocal learning. Making critical judgements about the performance of their teacher peers was useful practice for making judgements about the later performances of children. However, PA skills were not easily and automatically ac-

quired. Novices in assessment were likely to be insecure about their ability to assess and need explicit training to make reliable and acceptable assessment judgements.

## 5. Perception outcomes of PA

There were 21 studies of perceptions of PA, concerned with how the participant viewed the experience in both good and bad respects, usually in post hoc reflection. Some of these were completed immediately after the participants engaged in PA; most of them later. Most (n=13) obtained the views only of pre-service teachers, some obtained the views only of in-service teachers (n=4), two obtained the views of pre-service and in-service teachers at the same time and two obtained the views of in-service teachers and their school students at the same time.

### *Perceptions of pre-service teachers*

**Comparison of different methods over time:** Struyven, Dochy and Janssens (2008) investigated the effects of pre-service teachers' hands-on experience on their preferences for four assessment methods: portfolio, case-based, PA and multiple-choice evaluations. Data collection was by questionnaires, with a longitudinal design involving three moments of measurement. Results showed initially negative responses to unknown assessment methods. As familiarity with the assessment methods grew, preferences changed positively, as did perceptions of the appropriateness of the assessment method for evaluation purposes.

**Control and pre-post:** Only two studies out of 13 used a pre-post design. Sahin-Taskin (2018) used a pre-post design with a control group investigating primary pre-service teachers' pedagogical beliefs and self-efficacy, finding support for PA in relation to active learning practices. Bain (2009) explored the attitudes of design and technology pre-service teachers to PA prior and subsequent to a set of activities. Exposure altered participant's perceptions of PA, revealing a change from attitudes dominated by teacher-centred or didactic understandings to those where the role of learner was central.

**Post only after specific activity:** Özdemir (2016) surveyed 41 prospective Turkish teachers after they presented a lesson and were evaluated by groups of five peers. PA seemed beneficial for both academic and affective development, and in use of metacognitive strategies. Inexperience with PA and too many assessment criteria were problems, along with personalities interfering.

**Post only in general terms:** There were nine such studies. Al-Barakat and Al-Hassan (2009) investigated how PA enhanced 72 pre-service teachers' preparation during field experience. They had positive beliefs about PA. Bay (2011) administered a survey to 56 prospective English language teachers. They thought they acquired professional skills through PA and that PA provided constructive feedback in learning process and enhanced the quality of learning. However, PA was a tiring and time-consuming activity. Usage of rubrics, better comprehension of criteria and training were factors that improved the PA.

McGarr and Clifford (2013) reported on surveys and focus groups on PA in an 18-month pre-service programme. The vast majority of respondents felt the experience was valuable and enjoyable. However, many still felt teacher assessment was fairer and that peers should not have a say in their overall grade. Bratkovich (2014) investigated second language writing. Participants felt their writing ability had increased due to PA, which they preferred to self-assessment. Even when the feedback was not used, their writing ability increased, suggesting it nonetheless occasioned reflection. Mumm, Karm and Remmik (2016) investigated PA with 20 pre-service teachers. They said grade-centred summative assessment by teacher educators remained the principal way of assessing them, thus hindering the supportive role of other assessment for learning practices.

Cabello, Topping and Taylor (2017) explored how PA might enhance changes in 38 pre-service teachers' thinking about scientific concepts, and particularly the quality of instructional explanations. Questionnaires, feedback sessions and focus groups showed a transformation in thinking. Ratminingsih, Artini, Padmadewi and Ni (2017) used a questionnaire with 100 pre-service teachers and focus groups, who had very positive perceptions about PA, reporting it transformed their teaching. However, they were worried about subjectivity, felt under-confident and noted that PA took time.

Kurtuldu and Özkan (2019) used questionnaires and interviews with 29 pre-service English teachers involved in microteaching sessions. PA led to recognition of mistakes and improved skill levels, but there were concerns about the biasing effect of friendships (in both directions) and anxiety, and teacher assessment was thought more objective. Shahzadi and Hussain (2019) used a questionnaire with 1,374 pre-service teachers in public universities in Pakistan.

The majority agreed that PA helped them in identifying their weaknesses and strengths more effectively, but took a lot of time and had potential bias.

#### *Perceptions of pre-service and in-service teachers simultaneously*

Wen, Tsai, and Chang (2006) used a questionnaire with 280 pre-service teachers and 108 in-service teachers to evaluate attitudes towards PA. A difference was found between pre-service and in-service teachers, the latter viewing PA more as a learning aid. Male pre-service teachers had more positive attitudes towards PA. Thus, this study showed gender effects, which have otherwise been little studied. Brew, Riley, and Walta (2009) similarly surveyed both pre-service primary teaching students and the 30 teacher educators who taught them. The latter were far more supportive of PA than the former, but several teachers revealed that their high level of support for PA was closely linked to time-saving.

#### *Perceptions of in-service teachers*

Turning to perceptions of in-service teachers only, Woolhouse (1999) used PA with in-service teachers and found their reaction mostly positive. Similarly, Grob, Holmeier, and Labudde (2017) collected documentary and interview data from seven primary and ten secondary school teachers who actually implemented PA in their science classes. Teachers from both school levels believed that PA not only enhanced learning but also had social and motivational effects.

Less positively, Rotsaert, Panadero, and Schellens (2018) surveyed 225 secondary education teachers and found their perceived *accuracy* of PA was a major predictor of belief in its educational value. Teachers were not very aware of their school students' concerns about the impact of interpersonal processes in PA, or the importance attributed to anonymity. The amount PA was actually used in the classroom was investigated by Kippers, Wolterinck, Schildkamp, Poortman and Visscher (2018). Teachers reported they conducted PA in only 10%–25% of their lessons, while using direct instruction in 25%-50%.

#### *Perceptions of in-service teachers and school students*

Two studies investigated this. Jerotijević-Tišma (2016) explored attitudes to PA of English as a Foreign Language (EFL) of 112 teachers and 130 primary and secondary school students. There were generally positive perceptions of PA among both teachers and school students and this emphasized the importance of PA as a reliable classroom assessment practice. More specifically, Fidalgo and Thormann (2019) asked school students and their teachers for their opinions about PA in the context of e-Portfolios for learning information literacy skills. Most school students and all teachers thought that PA was a good way to teach such skills. PA was considered beneficial, but had some challenges.

## 6. Pre-Service outcomes for PA in TET

There were 11 studies of PA before 2009 and 24 after 2009, indicating the rise in interest in this area. The quality of studies also improved after 2009—all the control group studies (n=5) were in the later period.

**Control group studies:** Beaver and Scott (2011) studied peer grading with 58 elementary and middle school pre-service teachers. The experimental group did peer grading; the control group did not. Statistically significant positive changes were noted in a variety of the experimental groups' perceptions, in particular that PA could help them increase their own depth of knowledge about the topic. Strijbos, Narciss and Dünnebier (2010) researched 89 mixed pre-service teacher training and psychology participants, mostly female, assigned to four experimental and one control group. Those in Concise General Feedback groups outperformed Elaborated Specific Feedback groups. Elaborated Feedback by a high competent peer was perceived as more adequate, but led to more negative affect. Groups with low competence peers outperformed groups with high competence peers.

Birjandi and Hadidi Tamjid (2012) explored PA in promoting writing performance in 157 TEFL (Teaching English as Foreign Language) pre-service teachers, assigned to four experimental groups engaged in PA and self-assessment and a control group. On a pre-post writing test, both self-assessment and PA groups showed improvement in writing. Cabello (2013) used a peer assessment questionnaire, feedback sessions, focus groups and interviews in a ten-session PA with 37 pre-service science teachers in three Chilean universities, with an experimental and control group. The quality of their scientific explanations was measured at pre, post and follow-up in their eventual first job via observational analysis of video-recorded microteaching episodes. Results showed they significantly improved their explanations of scientific concepts, although not all were transferred into real teaching contexts. It was concluded that PA could play a noteworthy role in developing skills.



Weldeana and Sbhatu (2017) worked with 40 pre-service teachers in geometry, divided into treatment and comparison groups. The treatment group had Portfolios of Evidence, involving peer and self-assessment, and scored higher than controls on a mathematical test measuring Skills, Properties, Uses, and Representations. The treatment teachers made more statistically significant shifts towards supportive learning behaviours and abandoning inhibiting behaviours.

**PA after a specific activity:** Kilic and Cakan (2007) used PA with pre-service primary science teachers, who taught a science topic as a team to their peers, who then assessed their performance. Peer scores were considerably higher than teacher educator scores, but significantly correlated with them. Ozogul, Olina, and Sullivan (2008) randomly assigned preservice teachers to one of three treatment conditions: teacher-evaluation, self-evaluation or peer-evaluation. Three periods of instruction on writing lesson plans were held, then each participant submitted a draft lesson plan. These were assessed, with scores given and feedback provided on a 15-item rubric. All three treatment groups improved their lesson plans from draft version to final version, with the teacher-evaluation group showing significantly greater improvement. This is one of the few papers which did not find improvement for the PA group.

Iqbal and Mahmood (2008) involved 25 pre-service teachers in PA, observing their peers during teaching practice. This is the only study of observation during teaching practice in this review. Each pre-service teacher was observed by 22 class fellows and supervising teacher at least once. Five pre-service teachers with the highest and six with the lowest grades were analysed. The PA of the high-grade pre-service teachers was relatively similar to that of the teacher (seven out of 10 constructs) while low-grade teachers were less similar (four out of 10 constructs). There was more agreement on constructs involving tangible characteristics.

Li, Liu and Steckelberg (2010) investigated the relationship between quality of PA and quality of projects in a technology course for 43 pre-service teachers, who anonymously rated and commented on two randomly assigned peer projects then assessed the projects improved by feedback. Two independent raters then blindly evaluated initial and final projects. There was a significant relationship between the quality of peer feedback *provided* and the quality of their own final projects. However, there was no relationship between the quality of peer feedback *received* and their own projects.

Haynes and Miller (2015) used PA with 44 primary pre-service teachers focusing on motor skills. Performance in three trials at each of seven testing stations (for the run, strike, catch, throw, dodge, vertical jump and stationary kick) was assessed and simultaneously video-recorded for later re-assessment. The run and the overarm throw were selected for reporting. The pre-service teachers reported a high level of satisfaction regarding the value of PA as an awareness-raising exercise and as a tool to assist them to learn to analyse movement. Kiliç (2016) compared PA with self-assessments and teacher assessments. PA of 15 pre-service teachers' presentations was seen by participants as useful practice for developing more successful performance, higher confidence, effective presenting skills and essential competencies required for effective teaching.

Tsai (2012) interviewed 40 pre-service teachers who had experienced a PA activity. Each had to complete a research proposal and submit it to PA, reviewing peers' work and obtaining peer comments. Sixty percent of participants developed views of "epistemic relativism" (recognition of the diversity of knowledge perspectives involved in a research issue) and/or "social relativism" (an understanding of the multiplicity of peer perspectives). Their epistemological beliefs were related to their views of "epistemic relativism", while their Internet self-efficacy was associated with their views of "social relativism".

Çevik, Haşlamam and Serkan (2015) used PA with 82 pre-service teachers in complex problem solving. Both feedback and non-feedback conditions improved in developing solutions for the problems. Additionally, while feedback function and feedback direction predicted the use of feedback, agreement with feedback was not related to feedback use.

**Assessment criteria defined by students:** There were three studies in which assessment criteria had been designed by the students. Seery, Canty and Phelan (2012) used PA with 137 pre-service technology teachers, who defined the assessment criteria. There was a moderate to strong relationship between PA and teacher scores. Level of engagement, diversity and problem solving were identified as significant results of PA. Galichkina (2016) worked with ESL pre-service teachers, using project-based learning. She had participants develop assessment criteria for self- and peer-reflective weekly journals, developing descriptive rubrics for grading oral presentations. The trainees developed deeper assessment awareness of project-based learning activities, which helped develop confidence and gain process insights.

Tait-McCutcheon and Knewstubb (2018) investigated the alignment of self, peer and teacher assessment with 34 pre-service teachers. Peer groups assessed a teaching artefact using a collaboratively constructed assessment rubric and feedback sheet. While 60% of self, peer and lecturer-assessments were fully aligned, in the remaining 40% many of the

participants awarded lower grades to their own achievement, indicating less confidence than their peers or lecturer showed.

**Other interesting points:** A number of individual studies ( $n=7$ ) raised interesting points which were not discussed in any other paper. For example, Johnston (1993) used PA in music education (specifically, conducting) with pre-service teachers. They felt multiple assessment (PA and teacher assessment) was effective (although it took more time) and the inclusion of PA gave a more accurate view of conducting skills. Kilic and Cakan (2006) applied PA twice in a semester with pre-service primary science teachers, to assess individual contributions of group members to group projects.

Ross and Welsh (2007) successfully gave core tasks of progressively increasing demand subject to PA, associated with five distinct learning milestones. Li, Liu and Zhou (2012) examined the role of assessors and assessee in PA. The assessee's role was based on their ability to critically judge and act upon peer feedback, instead of simply the quality of peer feedback that they received. How participants responded to peer feedback significantly predicted their final project scores.

Oren (2012) examined the effects of gender in PA with pre-service science teachers. Females had significantly higher mean scores than males. Castejón, López-Pastor and Pérez-Pueyo (2012) investigated peer grading of exams in three pre-service teacher education programs. Results showed that correlations between grades given by pre-service teachers and teacher educators ranged from 0.66 to 0.90 (moderate to high) and although the average grades of the pre-service teacher group were lower than those of the teacher educators, they positively valued the experience.

Finally, Mewald (2016) deployed PA in the context of three pre-service teacher courses: young adult literature, methodology and language skills. PA in process writing brought about better texts but not necessarily teachers who were proficient and stable in the long term in providing formative assessment with criterion-oriented scales.

Thus, these seven papers each raised an issue; about: future use of both methods of assessment, assessment of individual contributions to a group project, application of PA to successively harder tasks, the use of feedback by assessee, the effects of gender, applying PA to examinations and the generalisation of PA skills from one domain to another.

**Studies of a general nature:** There were 12 studies of a general nature, which will be mentioned briefly. As early as 1979, Dodds found that pre-service teachers could function effectively as behaviour change agents for their peers. Dart and Clarke (1991) found pre-service teacher locus of control and motivation improved with PA. Zevenbergen (2001) had pre-service mathematics teachers reviewing posters created by their peers and PA was found to be effective. Linge-fjard and Holmquist (2005) successfully involved preservice secondary mathematics teachers in PA, finding more positive attitudes and competencies in mathematical modelling.

Kennedy and Allan (2009) found there was a shift in primary pre-service teachers' own identity as teachers and in their understanding of the role and purpose of assessment as a result of PA. Lynch, McNamara and Seery (2012) incorporated PA in a project with 47 pre-service teachers, targeting critical thinking and metacognitive skills. The quality of reflections through peer feedback was high, despite participants' citing a preference for teacher feedback. PA led to higher quality learning outcomes and enhanced critical thinking skills. Nilsson (2013) had 24 primary science pre-service teachers use PA to develop Pedagogical Content Knowledge. PA was found to have potential for PCK development.

Bektas (2013) used PA with 49 mostly female pre-service teachers from two classes. One class performed better than the other, so classes can vary. Patton and Marty-Snyder (2014) used PA in a physical education pre-service teacher education program, to better prepare pre-service teachers to assess their own future school students. There was PA of management of school student behaviours, class procedures, feedback given and use of instructional time. Guidelines for structuring PA within field experiences were provided, such as context, objectives, training, selection and matching of participants.

Eather, Riley, Miller and Jones (2017) used PA with 36 participants in a pre-service teachers' physical education course. PA led to significant improvements in perceived teaching confidence and competence, and teaching self-efficacy. Kaya and Özkan (2019) had 21 pre-service language teachers first carry out PA for a midterm exam then self-assessment for a final exam. Pre-service teachers scored both their peers and themselves in a similar manner to the teacher. Data showed benefits of this process, but the participants favoured self-assessment over PA.

Seifert and Feliks (2019) aimed to explore 300 pre-service teachers' opinions on the contribution of self-assessment and anonymous PA to the quality of their assignments and improvement of their assessment skills. They significantly benefitted from both processes. Anonymous evaluation allowed them to overcome inhibitions in evaluating peers' works.

## 7. In-service outcomes for PA in TET

The number of studies on PA in teacher in-service education (5) is far fewer than those in pre-service teacher education (35). This doubtless reflects the fact that pre-service teachers are readily available and are more controllable as research subjects. Two studies explored underpinning strategies and three focused on PA in the classroom.

**Underpinning strategies:** Harris and Brown (2013) reported on PA in three classrooms. Four themes were identified (improvement, accountability, social interaction and accuracy). In-service teachers cited mainly improvement purposes for PA, but understanding of how to use PA for improvement and self-regulation purposes was limited. Gómez Sará (2016) investigated PA in the spontaneous interactive speaking in English of 14 in-service teachers with another first language. Video recordings and two PA forms evidenced the development of three enhancement strategies (willingness to improve, use of compensatory strategies, and construction of a personalized version of the corpus) and two detrimental traits (underassessment and dependency on the corpus).

**PA applied in classrooms:** Ploegh, Tillema and Segers (2009) conducted a survey. Fifty-six in-service teachers using PA in classrooms rated several quality criteria as relevant to their teaching practice. PA was seen as under the control of the in-service teacher, who organized and orchestrated the process, giving participants clear guidelines to appraise one another's work. Panadero and Brown (2017) explored beliefs about PA in a survey of 750 primary, secondary, and higher education teachers in Spain. Teachers reported occasional use of PA in their classrooms but positive experience of it. The vast majority did not use anonymous forms of PA and half of the teachers considered the school students were accurate when assessing peers. The self-reported frequency of using PA was strongly predicted by teacher experience of PA.

Ma, Xin and Du (2018) proposed a peer coaching-based personalized learning PA approach compared to expert guidance to help 20 in-service teachers identify their questions and needs and adapt their teaching plans based on peer feedback, so as to engage in in-depth learning and transfer of knowledge to their teaching practice. The post-test scores of the peer-coaching group on learning participation, learning design skills and in-practice teaching abilities were significantly higher than those of the expert-led group.

## 8. Design considerations for PA in TET in general

This section partially comprises discussion of 11 further individual papers (eight dealing with pre-service teachers) which were generally of good quality and contained much practical advice, followed by a list of pointers for implementation drawn from these and sections 4-7. It follows that none of them are new, but they are newly gathered here. While these pointers are evidence-based, all the evidence supporting them is not explicated here, or the paper would be far too long.

### *Individual papers*

**Pre-service teachers:** Taking studies of pre-service teachers first, Sluijsmans, Brand-Gruwel and van Merriënboer (2002) reported PA with 93 pre-service teachers in a *controlled* study, *training* them in defining performance criteria, giving feedback and writing assessment reports, with practice tasks embedded in the course. Experimental participants had higher end of course grades than controls and were more satisfied. In another *controlled* study, Sluijsmans, Brand-Gruwel, van Merriënboer and Martens (2004) trained 93 mostly female second-year primary pre-service teachers on defining performance criteria. They argued that PA involved: (1) *defining* assessment criteria; (2) *judging* the performance of a peer in strengths and weaknesses; and (3) providing constructive *feedback*. The experimental group scored significantly higher on the use of criteria, but did not surpass the control group on devising a subsequent lesson plan. Li and Gao (2016) researched 130 pre-service teachers, the experimental group (divided into low, average, and high achieving) using PA and the *control group* discussion. Low- and average-achieving participants showed significantly improved performance, but *PA had less impact on the high-achieving*.

Cebrián-Robles, Franco-Mariscal and Blanco-López (2018) focused on *argumentation* in science with 57 pre-service teachers in a *controlled* study. A six-session *training* programme involved identifying the elements of arguments in order to design assessment *rubrics*. PA was then used with and without rubrics. Experimentals did significantly better on pre-post argumentation than controls. However, at follow-up six months later the difference had reduced.

Sluijsmans, Brand-Gruwel, van Merriënboer and Bastiaens (2003) used a *longitudinal within-subjects design* with 110 mostly female first year pre-service primary teachers in mathematics over seven months. The use of assessment criteria increased over time, positive comments increased, negative comments decreased, average scores on reflection



papers increased markedly and fear of assessment decreased markedly. However, females found it difficult to criticize their friends and were more negative towards PA.

Fund (2010) had pre-service teachers present written reflections for peer feedback, evaluating usefulness over a three-year period. All participants improved, but science teachers did better than mathematics teachers. Some were “willing receiver”, others “willing donor”, and some both. Feedback should be *reinforcing* and *challenging, thorough* and *sincere*, and be *friendly, non-insulting* and *non-judgemental*. Headington (2014) explored how 100 pre-service teachers developed informal learning communities identified at three points during the course for discussion of assessment and feedback. *Trust* and *continuity* emerged as inter-related themes.

López-Pastor, Pérez-Pueyo, Barba and Lorente-Catalán (2016) gathered evidence on *rubrics* in assessing writing with pre-service teachers. There were significant differences in previous experiences of rubrics and prior knowledge of grading criteria led to better performance. They felt that rubrics were very valuable and they indicated high intention to use rubrics in the future.

**Inservice teachers:** Põldoja, Väljataga, Laanpere and Tammets (2012) described an assessment tool (DigiMina) used with in-service teachers allowing *authentic, reliable and valid* assessment. Magno (2012) describes the creation of a *rubric*—the “Peer Assistance and Review Form” (PARF)—focused on four areas: planning and preparation, classroom environment, instruction, and professional responsibility. It was tested with 183 higher education teachers, observed by two assessors in their class. Barnes and King (1990) reported that a teacher certification law in Maine in 1984 required that support teachers be provided with *training* in peer observation and assessment techniques—an early example of *mandatory PA*.

## Implementation pointers

### *Points for initial reflection*

1. Consider the *type* of PA to be used.
2. Consider the *gender balance* of your sample.
3. PA has effects for both assessees and assessors—make sure both *know* this.
4. Pre-service and in-service teachers may *differ* in response.
5. Is PA accepted by the salient “*authorities*” in education?
6. Consider how PA fits with *teacher-dominated* instruction.
7. PA *will not save time* in the short term, but might in the long term.
8. Take care with *comparisons* of peer and teacher assessment.
9. What *social* and *motivational* effects are there from PA?
10. PA can be used on very *practical* skills (e.g., motor, music, presentation skills).
11. Abandonment of inhibiting behaviours can be important—*unlearning*.
12. PA may be *understood differently* by pre-/in-service teachers cf. school students.
13. PA training has been *mandated by law* in places.
14. Consider using PA of *individual contributions to group projects*.
15. PA might *impact less on high achievers*.
16. *Integrate PA* in an existing course rather than drawing attention to it as unusual.

### *Be aware of the whole process*

1. Take time to *ensure acceptance*.
2. Have participants consider *objectives* of PA—both *short-term* and *long-term*.
3. Have participants consider the *context* of application.
4. Participants’ *early negativity* often becomes *positive* over time with experience.

### *Selection and matching*

1. *Sustain* matching or *change* regularly or *give option*?
2. Is *anonymity* important to participants—at the *beginning* or *later*?
3. Pair friends or *avoid friends* or give choice?
4. *Social network analysis* might be useful in the matching process.
5. Will online contact be *synchronous* or *non-synchronous*?
6. Manage any *interpersonal issues* that arise.

### *Assessment criteria and training*

1. Use of criteria can *increase over time*.
2. *Co-develop assessment criteria* with participants.

3. Use a reasonably *short assessment form or rubric* with *clear* criteria.
4. Use a combination of both *tangible and more judgemental criteria* in a rubric.
5. *Rubrics* for PA in activities as unusual as conducting have been reported.
6. In complex areas, *identification of components* is necessary before rubric construction.
7. *Peer grading* or elaborated *formative feedback* or both?
8. Participants need explicit *training* in setting objectives, defining criteria, analysing performance and onward use of PA.
9. Provide explicit *practice* on non-threatening tasks.
10. Should there be *observational and/or video analysis* of microteaching or performance sessions? E.g. from Youtube.
11. Consider the use of *Facebook*.
12. To what extent should PA focus on *mistakes*?
13. Make clear if *explicit error correction* is expected.

### ***Confidence and self-efficacy***

1. Directly address participants' need for *confidence and trust*, particularly in *reliability*.
2. *Fear* should *decrease*—if not, discuss.
3. Make practical tasks *challenging*.
4. Have participants improve their *criticality* and *reliability*.
5. Create a stepwise progression of *increasingly demanding PA tasks*.
6. Consider peer assessment by *observation* of teaching skills.

### ***Feedback***

1. Participants need *training* in providing feedback.
2. Give guidance on amount of *positive* and *negative* feedback to be given.
3. Feedback should be both *reinforcing* and *challenging*.
4. Give instructional *support* (e.g. hyperlinks to multiple perspectives/conceptual knowledge).
5. Assesseees need to *reflect* and *select* from peer feedback what they will incorporate.
6. *Providing* feedback might have more impact on attainment than receiving it.
7. *Elaborated* feedback by a *highly competent* peer might not lead to positive effects.
8. *Simpler* feedback from *lower competence* participants might have better effects.
9. Feedback *enhances reflection* even when not accepted.
10. Some participants might be better at being *assessor* than *assessee*, or vice versa.
11. Do not assume revision will (only) *use* peer feedback.

### ***Stress from time and effort***

1. PA can be *tiring* and *time-consuming* and *somewhat unreliable*.
2. PA might not be the *favourite* form of assessment irrespective of effects.
3. How *aware* are teachers of school student personal factors?
4. Greater reliability might take *more time*.
5. Participants might *prefer* teacher feedback despite the positive effects of PA.
6. Evaluation results need to be *fed back* to the participants.
7. Sometimes participant assigned grades are *higher* than those of teachers.
8. PA *contributing to final grades* might be difficult for participants to accept.

### ***Measures***

1. Are you interested only in *outcomes* or only in *process* or both?
2. PA might be effective *despite* participant perceptions.
3. Consider measuring how participants perceive the *purpose* of assessment.
4. How to measure the impact of PA on *engagement, critical thinking, metacognition, complex problem solving*?
5. PA can develop realisation that different colleagues have *different points of view*.
6. PA can develop knowledge of *diversity of research perspectives*.
7. PA can affect the *quality of explanations* given.
8. PA effects might be specific and *not generalise*.

### ***Follow-up***

1. Ensure follow-up to check if gains have been *sustained*.
2. Build in carry-over to *using PA in the classroom* with school students.

3. Ensure transfer beyond specific PA tasks—build in *steps to self-regulation*.
4. Consider long term effects over *changes of context*.
5. Use of PA is a strong *predictor* of liking it.
6. *How much* PA should be used in school classes?
7. Consider articulation with self-assessment? Self-assessment after PA?

## C. Limitations, summary and conclusions

### 9. Critique of design and methodology of studies reviewed

The quality of some studies left a good deal to be desired, especially in Perceptions, although some good quality studies were found in Outcomes and Design. However, at this stage of development, it was felt more appropriate to conduct a review including many papers, rather than a best evidence synthesis or meta-analysis which would exclude a large number of papers. These latter can come later as the field develops and much higher quality papers are available.

Many of the studies reviewed here tended to have multiple problems, with design, sampling, measures, and analysis. The type of PA was not always made clear, and studies of peer grading and elaborated formative feedback were sometimes confused. Many studies did not have a control or alternative treatment group or a pre-post design. PA was found in some cases to improve over time with experience after initial reservations, and of course studies with only one moment of measurement were unlikely to discover this. Very few studies included follow-up, especially not when pre-service teachers were subsequently active in the classroom. Where samples represented the full cohort of participants in a course, this was acceptable, but we were rarely told of attrition rates owing to refusal, non-attendance or non-cooperation. However, samples often did not represent a full cohort, but instead were self-selected volunteers. This was clearly likely to bias the findings, probably positively.

Measures were also a problem. The Perceptions section detailed only the subjective perceptions of participants. In the other sections, more objective indicators of performance were used, but rarely more than one. Good examples were the use of final course grades by Sluijsmans, Brand-Gruwel and van Merriënboer (2002), the measurement of quality of pre-service teachers' scientific explanations at pre, post and follow-up in their eventual first job by Cabello (2013), and improved pre-post argumentation competence by Cebrián-Robles, Franco-Mariscal and Blanco-López (2018).

Analysis was also a problem. Many studies asserted positive findings for PA, but also commented on participants' resistance or dislike. Clearly, the picture was not all rosy. Where statistical analysis was used, it was often rather primitive. Additionally, with larger samples a finding of statistical significance could be misleading and very few studies gave effect sizes.

Most studies were of short-term interventions, although a few did investigate over three years of a pre-service teacher course. Very few included any follow-up, and very few studies explored the extent to which PA in training (even in-service training) carried through into PA with school students in the classroom. For the future, better quality studies with improvements in design, sampling, measurement, analysis and follow-through are needed.

### 10. Summary and conclusions

The search uncovered 73 papers of relevance. Underpinning theory was conspicuous by its absence (only one paper). There were 21 studies of Perception Outcomes of PA, using questionnaires, focus groups and interviews. Thirteen reported outcomes for pre-service teachers, four reported outcomes for in-service teachers, two reported outcomes for both, and two reported outcomes for in-service teachers and school students. Only two of these reported negative outcomes. However, six papers noted some issues and concerns despite this general positivity. Complaints included that PA was tiring and time-consuming, assessment criteria could be too numerous, personal factors might intervene and PA could be unreliable.

Turning to more objective outcomes from PA, there were 35 papers, of which 30 reported positive outcomes, only one reported a negative outcome, and in another four cases the positive/negative balance was unclear, in two cases because focus was entirely on comparison with teacher assessment. Studies showed positive gains on lesson plans, writing tests, mathematical tests, quality of explanations, developing assessment criteria for reflective journals, questions and descriptive rubrics for grading oral presentations and blog content, for example. Females tended to do better than males. There were only five papers on PA in in-service teacher education and training, and three of these were positive, with the positive/negative balance unclear in the other two cases.

The review then explored Design Considerations for PA. Eleven further studies were discussed with particular design implications, and evidence-based pointers for successful implementation drawn from these and the preceding sections.

The methodology and design of the studies was then critiqued. There were problems in design, sampling, measuring,

analysis and follow-through, especially into the teacher's own classroom. Studies almost always found that PA was effective and well received, but resistance at the beginning was noted quite frequently. However, studies with a control group or pre-post design were just as likely to find PA effective as other studies.

Overall, peer assessment seemed like a useful pedagogical tool. That does not mean it offered automatic saving of time, since to implement it thoroughly took a good deal of staff time. Given the mixed quality of these studies, caution in conclusions is needed. What is clear is that at the moment is that:

1. There is a serious lack of relevant theory on PA in TET.
2. Perceptions of PA are generally positive for both pre-service and in-service teachers.
3. On more objective outcomes many pre-service teachers show positive outcomes, while fewer in-service teachers are still mostly positive.
4. The outlining of Design Considerations should help with improved future PA projects and evaluation research.
5. There is almost no research on PA with pre-service and in-service teachers leading to initiating or improving PA for school students in classrooms.

## References

- Al-Barakat, A. and Al-Hassan, O. (2009). Peer assessment as a learning tool for enhancing student teachers' preparation. *Asia-Pacific Journal of Teacher Education*, 37(4), 399-413.
- Annis, L. F. (1983). The processes and effects of peer learning. *Human Learning*, 2(1), 39-47.
- Bain, J. (2009). Attitudes toward peer assessment in initial teacher education students: An exploratory case study. In: Norman, E., & Spendlove, D. (Eds.), *The Design and Technology Association International Research Conference*, Loughborough University, 30 June-2<sup>nd</sup> July 2009. Wellesbourne: The Design and Technology Association.
- Barnes, K. J. and King, D. R. (1990). Peer observation and assessment inservice training: Focusing on support teams and teacher action plans. Paper presented at the Annual Meeting of the Eastern Educational Research Association (Clearwater, FL, February, 1990). ERIC Document Reproduction Service ED324307.
- Bay, E. (2011). The opinions of prospective teachers about peer assessment. *Gaziantep University Journal of Social Sciences*, 10(2), 909-925.
- Beaver, C. and Beaver, S. (2011). The effect of peer-assessment on the attitudes of pre-service elementary and middle school teachers about writing and assessing mathematics. *Issues in the Undergraduate Mathematics Preparation of School Teachers*, 5. Available at [www.k-12prep.math.ttu.edu](http://www.k-12prep.math.ttu.edu). ERIC Document Reproduction Service Number: EJ962630.
- Bektas, M. (2013). Teaching practice of life study lesson of classroom teacher candidates' analysis of the results of peer assessment. *Educational Research and Reviews*, 8(17), 1621-1625.
- Birjandi, P. and Hadidi Tamjid, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education*, 37(5), 513-533.
- Bratkovich, M. O. (2014). Assessment and feedback: Examining the relationship between self-assessment and blind peer- and teacher-assessment in TOEFL writing. *Working Papers in TESOL & Applied Linguistics*, 14(2), 100-127.
- Brew, C., Riley, P., and Walta, C. (2009). Education students and their teachers: Comparing views on participative assessment practices. *Assessment & Evaluation in Higher Education*, 34(6), 641-657.
- Cabello, V. M. (2013). Developing skills to explain scientific concepts during initial teacher education: The role of peer assessment. Doctoral thesis, University of Dundee. Available at <https://www.researchgate.net/publication/312540075>.
- Cabello, V. M., Topping, K. J., and Taylor, N. (2017). Formative peer assessment during initial teacher education: An experience of changing conceptions about explanations. *Revista Sul Americana de Psicologia*, 5(1), 1-25.
- Castejón, J., López-Pastor, V. M., and Pérez-Pueyo, A. (2012). Involving students in teacher education assessment? A case study of a peer evaluation experience. *REMIE: Multidisciplinary Journal of Educational Research*, 2(2), 177-201.
- Cebrián-Robles, D., Franco-Mariscal, A-J., and Blanco-López, Á. (2018). Preservice elementary science teachers' argumentation competence: Impact of a training programme. *Instructional Science: An International Journal of the Learning Sciences*, 46(5), 789-817.
- Çevik, Y. D., Haşlamam, T., and Serkan, Ç. (2015). The effect of peer assessment on problem solving skills of prospective teachers supported by online learning activities. *Studies in Educational Evaluation*, 44, 23-35.
- Dart, B. C. and Clarke, J. A. (1991). Helping students become better learners: A case study in teacher education. *Higher Education*, 22(3), 317-335.
- Dodds, P. (1979). A peer assessment model for student teacher supervision. *Research Quarterly*, 50(1), 18-29.

- Eather, N., Riley, N., Miller, D., and Jones, B. (2017). Evaluating the effectiveness of using Peer-Dialogue Assessment (PDA) for improving pre-service teachers' perceived confidence and competence to teach physical education. *Australian Journal of Teacher Education*, 42(1), 69-83.
- Falchikov, N. and Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322.
- Fidalgo, P. and Thormann, J. (2019). Student and instructor opinions about using eportfolios to learn information literacy skills. *Journal of Educational Multimedia and Hypermedia*, 28(2), 165-191.
- Fund, Z. (2010). Effects of communities of reflecting peers on student-teacher development—including in-depth case studies. *Teachers and Teaching: Theory and Practice*, 16(6), 679-701.
- Galichkina, E. (2016). Developing teacher-trainees' assessment awareness in the EFL classroom through project-based learning activity. *Journal of Language and Education*, 2(3), 54-60. DOI: 10.17323/2411-7390-2016-2-3-54-60. Available at SSRN: <https://ssrn.com/abstract=3060643>.
- Gielen, S., Dochy, F., and Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education*, 36(2), 137-155. DOI: 10.1080/02602930903221444.
- Gómez Sará, M. M. (2016). The influence of peer assessment and the use of corpus for the development of speaking skills in in-service teachers. *HOW*, 23(1), 103-128. <http://dx.doi.org/10.19183/how.23.1.142>.
- Grob, R., Holmeier, M., and Labudde, P. (2017). *A teacher perspective on benefits and challenges of peer-assessment*. Paper presented at ESERA17 conference, 21<sup>st</sup>-25<sup>th</sup> August 2017, Dublin City University, Dublin, Ireland.
- Harris, L. R. and Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101-111.
- Hattie, J. and Wollenschlager, M. (2014). A conceptualization of feedback. In H. Ditton and A. Muller (Eds.), *Feedback und Rückmeldungen: Theoretische Grundlagen, empirische Befunde, praktischer Anwendungsfelder* (pp. 135-149). Munster & New York: Waxmann.
- Haynes, J. and Miller, J. (2015). Preparing pre-service primary school teachers to assess fundamental motor skills: Two skills and two approaches. *Physical Education and Sport Pedagogy*, 20(4), 397-408.
- Headington, R. (2014). *Undergraduate initial teacher education students' peer assessment and feedback communities: A longitudinal study*. Paper given at EARLI Assessment SIG Conference, Madrid, Spain.
- Hodgson, C. (2010). Assessment for learning in science: What works? *Primary Science*, 115, 14-16.
- Hoogeveen, M. and van Gelderen, A. (2013). What works in writing with peer response? A review of intervention studies with children and adolescents. *Educational Psychology Review*, 25(4), 473-502. DOI: 10.1007/s10648-013-9229-z.
- Iqbal, Z. and Mahmood, N. (2008). Compatibility of peer assessment and teacher assessment in observational situations: An emerging assessment tool in higher education. *Bulletin of Education and Research*, 30(2), 61-77. Available at SSRN: <https://ssrn.com/abstract=2362308>.
- Jerotijević-Tišma, D. M. (2016). Serbian EFL teachers and learners' attitudes to peer assessment and the comparison of teacher and peer speaking assessment scores. *Nastavai Vaspitanje*, 65(2), 379-393. DOI: 10.5937/nasvas1602379j.
- Johnston, H. (1993). The use of video self-assessment, peer-assessment, and instructor feedback in evaluating conducting skills in music student teachers. *British Journal of Music Education*, 10(1), 57-63. DOI: <https://doi.org/10.1017/S0265051700001431>.
- Kane, J. S. and Lawler, E. E. (1978). Methods of peer assessment. *Psychological Bulletin*, 85(3), 555-586.
- Kaya, D. and Özkan, Y. (2019). Using alternative assessment to engage preservice language teachers in the assessment process: A case study in Turkish higher education context. *Reading Matrix: An International Online Journal*, 19(1), 93-113.
- Kennedy, A. and Allan, J. (2009). 'The assessor and the assessed': Learning from students' reflections on peer evaluation. *Journal of Teacher Education and Teachers' Work*, 1(1), 4-17.
- Kiliç, D. (2016). An examination of using self-, peer-, and teacher-assessment in higher education: A case study in teacher education. *Higher Education Studies*, 6(1), 136-144.
- Kilic, G. B. and Cakan, M. (2006). The analysis of the impact of individual weighting factor on individual scores. *Assessment & Evaluation in Higher Education*, 31(6), 639-654.
- Kilic, G. B. and Cakan, M. (2007). Peer assessment of elementary science teaching skills. *Journal of Science Teacher Education*, 18(1), 91-107.
- Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., and Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, 75, 199-213.
- Kurtuldu, E. and Özkan, Y. (2019). Pre-service language teachers' reflections on peer assessment in micro teaching sessions of a methodology course. *Usak University Journal of Social Sciences*, December, 276-284.



- Li, L., Liu, X., and Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41(3), 525-536.
- Li, L., Liu, L. X., and Zhou, Y. (2012). Give and take: A re-analysis of assessor and assessee's roles in technology-facilitated peer assessment. *British Journal of Educational Technology*, 43(3), 376-384.
- Li, L. and Gao, F. (2016). The effect of peer assessment on project performance of students at different learning levels. *Assessment & Evaluation in Higher Education*, 41(6), 885-900.
- Li, H., Xiong, Y., Zang, X., Kornhaber, M. L., Lyu, Y., Chung, K. S., and Suen, H. K. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245-264.
- Lingefjard, T. and Holmquist, M. (2005). To assess students' attitudes, skills and competencies in mathematical modeling. *Teaching Mathematics and Its Applications: An International Journal of the IMA*, 24(2-3), 123-133.
- López-Pastor, V. M., Pérez-Pueyo, A., Barba, J. J., and Lorente-Catalán, E. (2016). Students' perceptions of a graduated scale used for self-assessment and peer-assessment of written work in pre-service physical education teacher education (PETE). *Ciencia*, 11, 37-50.
- Lynch, R., McNamara, P. M., and Seery, N. (2012). Promoting deep learning in a teacher education programme through self- and peer-assessment and feedback. *European Journal of Teacher Education*, 35(2), 179-197.
- Ma, N., Xin, S., and Du, J. Y. (2018). A peer coaching-based professional development approach to improving the learning participation and learning design skills of in-service teachers. *Journal of Educational Technology & Society*, 21(2), 291-304.
- McGarr, O. and Clifford, A. M. (2013). "Just enough to make you take it seriously": Exploring students' attitudes towards peer assessment. *Higher Education: The International Journal of Higher Education and Educational Planning*, 65(6), 677-693.
- Magno, C. (2012). Assessing higher education teachers through peer assistance and review. *The International Journal of Educational and Psychological Assessment*, 9(2), Available at SSRN: <https://ssrn.com/abstract=2287173>.
- Mewald, C. (2016). Process Writing and peer-assessment in teacher education. In Tsagari, D. (Ed.), *Classroom-based Assessment in L2 Contexts*. (pp.229-253). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Mumm, K., Karm, M., and Remmik, M. (2016). Assessment for learning: Why assessment does not always support student teachers' learning. *Journal of Further and Higher Education*, 40(6), 780-803.
- Nilsson, P. (2013). What do we know and where do we go? Formative assessment in developing student teachers' professional learning of teaching science. *Teachers and Teaching: Theory and Practice*, 19(2), 188-201.
- Oren, F. S. (2012). The effects of gender and previous experience on the approach of self and peer assessment: A case from Turkey. *Innovations in Education and Teaching International*, 49(2), 123-133.
- Özdemir, S. (2016). The opinions of prospective teachers on peer assessment. *Educational Research and Reviews*, 11(20), 1859-1870.
- Ozogul, G., Olina, Z., and Sullivan, H. (2008). Teacher, self and peer evaluation of lesson plans written by preservice teachers. *Educational Technology Research and Development*, 56(2), 181-201.
- Panadero, E. and Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education*, 32(1), 133-156.
- Patton, B. J. and Marty-Snyder, M. (2014). Practical applications for using peer assessment in physical education teacher education field experiences. *Strategies: A Journal for Physical and Sport Educators*, 27(6), 25-31.
- Ploegh, K., Tillema, H. H., and Segers, M. S. R. (2009). In search of quality criteria in peer assessment practices. *Studies in Educational Evaluation*, 35(2-3), 102-109.
- Pöldoja, H., Väljataga, T., Laanpere, M., and Tammets, K. (2012). Web-based self- and peer-assessment of teachers' digital competencies. *World Wide Web*, 17(2), 255-269. DOI: 10.1007/s11280-012-0176-2.
- Ratminingsih, N. M., Artini, L. P., and Padmadewi, N. N. (2017). Incorporating self and peer assessment in reflective teaching practices. *International Journal of Instruction*, 10(4), 165-184.
- Ross, M. M. B. and Welsh, M. P. (2007). Formative feedback to improve learning on a teacher education degree using a personal learning environment. *International Journal of Emerging Technologies in Learning (iJET)*, 2(3).
- Rotsaert, T., Panadero, E., and Schellens, T. (2018). Peer assessment use, its social nature challenges and perceived educational value: A teachers' survey study. *Studies in Educational Evaluation*, 59, 124-132. <https://doi.org/10.1016/j.stueduc.2018.07.001>.
- Sahin-Taskin, C. (2018). Effects of active learning environments supported with self- and peer assessment on pre-service teachers' pedagogical and self-efficacy beliefs. *Asia-Pacific Journal of Teacher Education*, 46(5), 421-440.
- Seery, N., Canty, D., and Phelan, P. (2012). The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology and Design Education*, 22(2), 205-226.
- Seifert, T. and Feliks, O. (2019). Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills.

- Assessment & Evaluation in Higher Education*, 44(2), 169-185.
- Shahzadi, U. and Hussain, B. (2019). Perceptions of prospective teachers about peer assessment as a tool for reflective practices. *Review of Economics and Development Studies*, 5(2). DOI: 10.26710/reads.v5i2.623.
- Sluijsmans, D. and Prins, F. (2006). A conceptual framework for integrating peer assessment in teacher education. *Studies in Educational Evaluation*, 32(1), 6-22.
- Sluijsmans, D. M. A., Brand-Gruwel, S., and van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment & Evaluation in Higher Education*, 27(5), 443-454.
- Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., and Bastiaens, T. J. (2003). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation*, 29(1), 23-42.
- Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., and Martens, R. (2004). Training teachers in peer-assessment skills: Effects on performance and perceptions. *Innovations in Education and Teaching International*, 41(1), 59-78.
- Strijbos, J. W., Narciss, S., and Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291-303.
- Struyven, K., Dochy, F., and Janssens, S. (2008). The effects of hands-on experience on students' preferences for assessment methods. *Journal of Teacher Education*, 59(1), 69-88.
- Tait-McCutcheon, S. and Knewstubb, B. (2018). Evaluating the alignment of self, peer and lecture assessment in an Aotearoa New Zealand pre-service teacher education course. *Assessment & Evaluation in Higher Education*, 43(5), 772-785.
- Tillema, H., Leenknicht, M., and Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning—A review of research studies. *Studies in Educational Evaluation*, 37(1), 25-34.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.
- Topping, K. J. (2018). *Using peer assessment to inspire reflection and learning*. Student assessment for educators series (Ed. J. H. MacMillan). New York & London: Routledge. ISBN: 978-0-8153-6765-9 (pbk). [www.routledge.com/9780815367659](http://www.routledge.com/9780815367659).
- Tsai, C. C. (2012). The development of epistemic relativism versus social relativism via online peer assessment, and their relations with epistemological beliefs and internet self-efficacy. *Educational Technology & Society*, 15(2), 309-316.
- University of Glasgow Story. (2020). George Jardine. Retrieved from <https://www.universitystory.gla.ac.uk/biography/?id=WH2384&type=P>.
- Van Gennip, N. A. E., Segers, M., and Tillema, H. M. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, 4(1), 41-54.
- Weldeana, H. N. and Sbhata, D. B. (2017). Portfolio of evidence: An assessment tool in promoting geometry achievement among teacher education college students. *EURASIA Journal of Mathematics Science and Technology Education*, 13(6), 1981-2004.
- Wen, M. C. L., Tsai, C. C., and Chang, C. Y. (2006). Attitudes towards peer assessment: A comparison of the perspectives of pre-service and in-service teachers. *Innovations in Education and Teaching International*, 43(1), 83-92.
- Woolhouse, M. (1999). Peer assessment: The participants' perception of two activities on a further education teacher education course. *Journal of Further and Higher Education*, 23(2), 211-219. DOI: 10.1080/0309877990230205.
- Zevenbergen, R. (2001). Peer assessment of student constructed posters: Assessment alternatives in preservice mathematics education. *Journal of Mathematics Teacher Education*, 4(2), 95-113.